

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2001-331355

(43)Date of publication of application : 30.11.2001

(51)Int. CI.

G06F 12/00

G06F 3/06

(21)Application number : 2000-152672

(71)Applicant : HITACHI LTD

(22)Date of filing : 18.05.2000

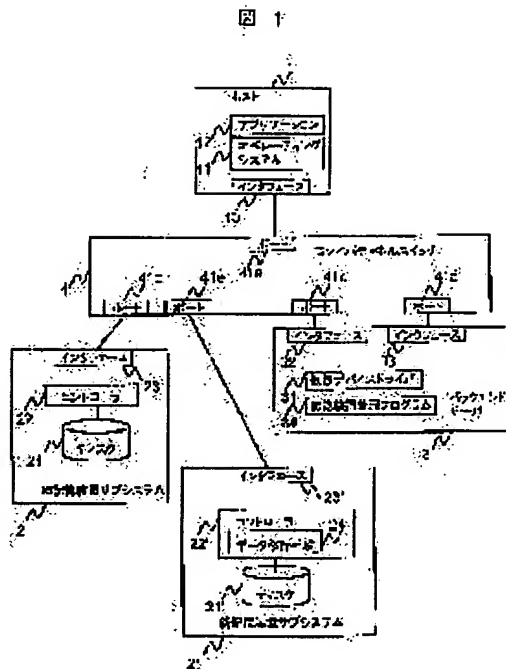
(72)Inventor : KITAMURA MANABU
ARAI HIROHARU

(54) COMPUTER SYSTEM

(57)Abstract:

PROBLEM TO BE SOLVED: To transmissively execute data transition among storage devices to host computers in a computer system, in which plural host computers are connected with plural storage devices.

SOLUTION: A back-end server 3 provides a host 1 having a virtual disk. The virtual disk first appears to be same as an old storage device sub-system 2 to the host 1. When data is transitioned from the old storage device sub-system 2 to a new storage device sub-system 2, the back-end server 3 first instructs a data transition processing to the new storage device sub-system 2, and after that, switches setting of the virtual disk and makes it correspond to the new storage device sub-system 2. Data transition among the disk devices is performed transmissively to the host 1, since this switching is transmissively executed to the host 1.



LEGAL STATUS

[Date of request for examination] 01.09.2003

[Date of sending the examiner's decision of rejection]

BEST AVAILABLE COPY

[Kind of final disposal of application
other than the examiner's decision of
rejection or application converted
registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's
decision of rejection]

[Date of requesting appeal against
examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998, 2003 Japan Patent Office

BEST AVAILABLE COPY

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2001-331355

(P2001-331355A)

(43) 公開日 平成13年11月30日(2001.11.30)

(51) Int.Cl. ⁷	識別記号	F I	データ* (参考)
G 0 6 F 12/00	5 1 4	G 0 6 F 12/00	5 1 4 5 B 0 6 J
	5 4 5		5 4 5 A 5 B 0 8 2
3/06	3 0 1	3/06	3 0 1 X
	3 0 4		3 0 4 F

審査請求 未請求 請求項の数 2 O L (全 7 頁)

(21) 出願番号 特願2000-152672(P2000-152672)

(22) 出願日 平成12年5月18日(2000.5.18)

(71) 出願人 000005108

株式会社日立製作所

東京都千代田区神田駿河台四丁目6番地

(72) 発明者 北村 学

神奈川県川崎市麻生区王禅寺1099番地 株

式会社日立製作所システム開発研究所内

(72) 発明者 荒井 弘治

神奈川県小田原市国府津2880番地 株式会

社日立製作所ストレージシステム事業部内

(74) 代理人 100076096

弁理士 作田 康夫

Fターム(参考) 5B065 BA01 CE22 EA33 EA35 ZA01

5B082 FA00 HA05

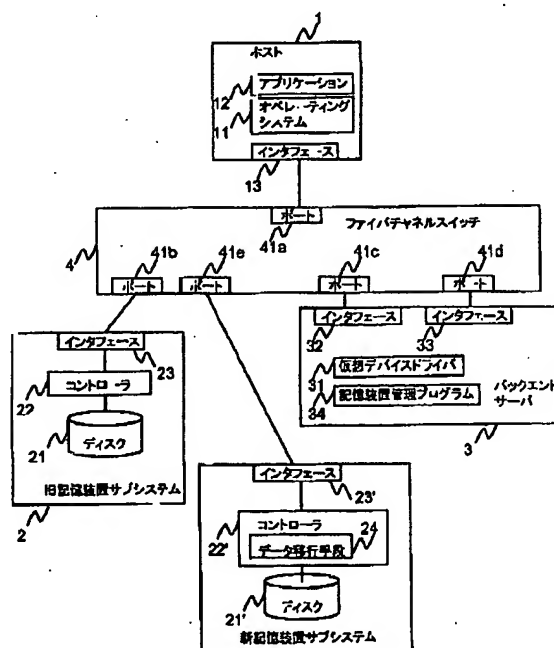
(54) 【発明の名称】 計算機システム

(57) 【要約】

【課題】複数のホストコンピュータと複数の記憶装置が相互結合された計算機システムにおいて、記憶装置間でのデータの移動をホストコンピュータに対して透過的に実施する。

【解決手段】バックエンドサーバ3はホスト1に対し、仮想ディスクを提供する。仮想ディスクははじめ、旧記憶装置サブシステム2と同じものとしてホスト1に見える。旧記憶装置サブシステム2から新記憶装置サブシステム2にデータを移行する場合、バックエンドサーバ3は最初に新記憶装置サブシステム2にデータ移行処理を指示し、引き続き仮想ディスクの設定を切り替えて新記憶装置サブシステム2に対応させる。この切り替えはホスト1に対して透過的に実施されるため、ホスト1に対して透過的にディスク装置間のデータ移行が可能になる。

図 1



BEST AVAILABLE COPY

【特許請求の範囲】

【請求項1】 複数の計算機と複数の記憶装置と、前記複数の計算機と複数の記憶装置とを相互に結合するスイッチとで構成された計算機システムにおいて、該計算機システムは前記複数の計算機に対し仮想的な記憶装置を提供する手段を有し、前記仮想的な記憶装置は前記複数の記憶装置の少なくとも1つの前記記憶装置に対応する記憶装置であって、前記仮想的な記憶装置を提供する手段は、前記対応を動的に変更することを特徴とする計算機システム。

【請求項2】 請求項1における、仮想的な記憶装置を提供する手段は、前記対応を動的に変更した際に、前記複数の計算機に対しては、前記対応が変化したことを見せないことを特徴とする計算機システム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、情報処理システムなどにおける記憶装置システムのデータアクセス方法に係り、特に、記憶装置内のデータ移行方法に関する。

【0002】

【従来の技術】パソコン、ワークステーション、メインフレームなどの異なるアーキテクチャ、オペレーティングシステムを採用しているプラットフォームと複数の記憶装置とを相互に接続し、いわゆる1つのネットワークにまとめる動きが盛んになっている。これを一般に、複数の計算機をイーサネット（登録商標）（Ethernet（登録商標））などのネットワークで接続したLAN（Local Area Network）に対する言葉でSAN（Storage Area Network）と呼ぶ。SANは通常ファイバチャネル（Fibre Channel）という光ケーブルないし銅線の伝送路を用いて計算機と記憶装置を接続する。

【0003】SANにはいくつかの利点があげられている。まず第1に複数の計算機から記憶装置が共通にアクセスできる環境を提供することである。第2に記憶装置同士も相互接続されることにより記憶装置間でのデータ転送が可能で、これにより、ホスト計算機に負荷をかけることなくバックアップや記憶装置間のデータコピーが実現でき、記憶装置障害時には副系の記憶装置への切り替えが可能となる。第3に、これまで個々の計算機に個々の記憶装置が接続されていたため、記憶装置の管理（装置の状態の監視、設定の変更）は接続されている個々の計算機からしかできなかったものを、特定の計算機から全ての記憶装置の管理を可能にする。また、従来のSCSI（Small Computer System Interface）では最高16台までの機器しか接続できなかったが、ファイバチャネルによって100台以上の機器をオンラインで接続でき、容易な拡張性を得られる。

【0004】近年、SANを実現するための製品が数多く現れてきているが、実際に上記利点を生かしたものはない。とくに拡張性においては、機器のオンライン接続は

物理的に可能になったものの、それを活用する基盤技術が不足している。たとえばSANにおいて、ディスク装置の交換のために新規にディスク装置を増設した場合、機器の増設はオンラインにて実施できるが、そのあとでデータの移動をユーザが明示的に行う必要がある。オンラインの機器増設でユーザがメリットを享受するには、単純なハードウェアの増設だけでなく、ハードウェアの増設に伴いデータ移動などがユーザに対して透過的に実施される必要がある。

【0005】ディスク装置間の、オンラインのデータの移動に関しては、米国特許5680640号にその例が開示されている。米国特許5680640号はメインフレーム用ディスクを前提としたデータ移行であるが、ディスク装置間を接続する通信線を利用し、ホストとディスク装置間の接続を短時間切断するだけで、あとはユーザに透過的にディスク装置間のデータ移行を可能にしている。

【0006】

【発明が解決しようとする課題】米国特許5680640号はユーザに対して限りなく透過的にディスク装置間でのデータ移動を可能にしている。ただし、これはメインフレーム用ディスクを前提としたデータ移行方法であり、SANにおいての適用は出来ない。米国特許5680640号では旧ディスク装置を新規ディスク装置に切り替える際、ディスク装置側の設定によって新規ディスクがあたかも旧ディスク装置であるかのようにホスト側に見せかけることが出来る。これはディスク装置のデバイス番号などの設定を操作することで可能である。

【0007】ただし、SAN、たとえばファイバチャネル環境の場合には、個々のディスクに付与される一意なIDは、ネットワークを構成する機器（ディスク装置、ファイバチャネルスイッチ）同士のネゴシエーションによって決定され、ユーザの設定によって変えられるものではない。米国特許5680640号のデータ移行方法を用いる場合、ホストコンピュータに対して、新規ディスク装置を旧ディスク装置として見せかけることはできず、事実上ユーザに透過的なデータ移行は実現できない。本発明の目的は、ホスト、ユーザに対して透過的で、かつSANの拡張性を生かすことのできるシステムを提供することにある。

【0008】

【課題を解決するための手段】本発明における計算機システムは、ホスト計算機、バックエンド計算機、複数の記憶装置サブシステムと、ホスト計算機とバックエンド計算機とを接続するスイッチとで構成される。ホスト計算機はバックエンド計算機を介して各記憶装置サブシステムにアクセスするが、バックエンド計算機は、ホスト計算機に対して1つないし複数の仮想的なディスク装置を提供する。ホスト計算機から仮想的なディスク装置にアクセス要求があると、バックエンド計算機では要求のあった仮想的なディスク装置の種類に応じて、実際に接

続されている記憶装置サブシステムに適宜要求を出す。

【0009】

【発明の実施の形態】図1は、本発明を適用した計算機システムの一実施形態における構成例を示すブロック図である。計算機システムは、ホスト1、旧記憶装置サブシステム2、新記憶装置サブシステム2、バックエンドサーバ3、ファイバチャネルスイッチ4とで構成される。

【0010】ホスト1はオペレーティングシステム11、アプリケーション12、インタフェース13から構成される。オペレーティングシステム11、アプリケーション12は実際にはホスト1上のCPU、メモリ上で動作するが、これらハードウェアの構成要素については本発明の内容と関係が無いため省略している。実際にはホスト1以外に複数のホストコンピュータがつながる環境が一般的であるが、本発明では簡単のため、ホストコンピュータとしてホスト1のみを記載している。旧記憶装置サブシステム2はディスク21、コントローラ22、インタフェース23とから構成される。ディスク21は複数の物理ディスクをまとめて1つの論理的なディスク装置に見せかけた論理ドライブであっても、本発明の内容に変わりはない。インタフェース23はファイバチャネルスイッチ4と接続される。新記憶装置サブシステム2も旧記憶装置サブシステム2と同様、ディスク21、コントローラ22、インタフェース23とから構成される。旧記憶装置サブシステム2と異なる点は、コントローラ22中にデータ移行手段24が含まれることである。

【0011】バックエンドサーバ3は仮想デバイスドライバ31、インタフェース32、33から構成される。仮想デバイスドライバ31はバックエンドサーバ3上のCPU、メモリ上で動作するソフトウェアで、ユーザによって外部から設定を変更したりあるいはプログラム自体の入れ替えをすることが可能であるが、CPU、メモリなどのハードウェア構成要素に関しては本発明の内容と関係ないため省略している。

【0012】ファイバチャネルスイッチ4は複数のポート41a、41b、41c、41d、41e（以下総称してポート41と略す）から構成され、ホスト1、旧記憶装置サブシステム2、新記憶装置サブシステム2、バックエンドサーバ3を相互に接続するために使用される。ポート41aからはいずれもポート41b、41c、41d、41eにアクセスすることが可能である。そのため、ホスト1はポート41b、41eから直接旧記憶装置サブシステム2や新記憶装置サブシステム2にアクセスすることもできるが、本実施形態においては、基本的にホスト1はすべてバックエンドサーバ3を介して記憶装置サブシステム2にアクセスすることとする。

【0013】バックエンドサーバ3の役割について説明する。バックエンドサーバ3は仮想デバイスドライバ3

1によって、ホスト1から見てあたかも1つないし複数のディスク装置であるかのように見える。本実施形態では、ホスト1がポート41dを介してインタフェース33を見ると、1つのディスクがつながっているように見えるものとする。以降、このディスクのことを仮想ディスクと呼ぶ。仮想デバイスドライバ31は、最初はホスト1からは仮想ディスクが旧記憶装置サブシステム2のディスク21と同じ物に見えるように設定されている。すなわちホスト1が仮想ディスクの論理ブロックアドレス(LBA)0あるいはLBA1にアクセスすると、仮想デバイスドライバ31はインタフェース32、ポート41cを介して、ディスク21のLBA0あるいはLBA1にアクセスし、結果をインタフェース33、ポート41dを介してホスト1に返す。本発明の実施形態ではインタフェース32がディスク21やディスク21にアクセスするために使われ、またインタフェース33がホスト1とのやり取りに使われるようになっているが、1つのインタフェースでこれら2つの役割を行わせることも可能である。また、仮想デバイスドライバ31の設定を変えることで、仮想ディスクが新記憶装置サブシステム2のディスク21に見えるようにすることも可能である。設定変更を行った場合、ホストコンピュータから見える仮想ディスクに変化はない。ファイバチャネルインタフェースを持つディスク装置の場合、ホストコンピュータからはポートIDと論理ユニット番号(LUN)で一意にディスク装置が認識できるが、仮想デバイスドライバの設定を変更して仮想ディスクがディスク21からディスク21に変更されたとしても、ホスト1に対して見える仮想ディスクのポートIDとLUNは変化せず、ホスト1は実際にアクセスしているディスクが変わったことの認識はない。次に新記憶装置サブシステム2のデータ移行手段24について説明する。データ移行手段24は米国特許5680640号に開示されているものと同様の手段を有する。データの移行が指示されると、データ移行手段24は記憶装置サブシステム2のディスク21の先頭から順にデータを読み出し、ディスク21へとデータを書き込む。さらに各ブロックないしは複数ブロック単位に、データの移行が終了したかどうかを記録するテーブルを持ち、移行処理中にリードアクセスがくると、このテーブルを参照し、データ移行が済んでいない領域についてはディスク21からデータを読み出し、データ移行が済んでいる領域についてはディスク21のデータを返す。

【0014】図2はデータ移行手段のもつテーブル100を表したものである。データ移行手段24ではブロックごとにデータをディスク21からディスク21へとコピーしていく。テーブル100ではそれぞれのアドレス101ごとにフラグ102を持つ。フラグ102が1である場合には、そのアドレスのデータはすでにディスク21からディスク21にコピーされたことを示し、0の場合には未コピーであることを示す。データ移行処理

BEST AVAILABLE COPY

や、データ移行処理中のリード、ライト処理ではこのテーブル100を利用する。

【0015】図3で、データ移行手段24の行う移行処理の流れを説明する。まずカウンタBを用意し、初期値を0とする(ステップ2001)。次にテーブル100を参照し、LBA Bのフラグ102が1かどうかチェックする(ステップ2002)。フラグが1の場合にはデータ移行が済んでいるため、カウンタBを1増加する(ステップ2005)。また、ステップ2002で、フラグ102が0であれば、ディスク21からディスク21へとデータをコピーし(ステップ2003)、テーブル100の該当するフラグ102を1に更新し(ステップ2004)、ステップ2005へと進む。ステップ2006ではディスク21の最終LBAまで処理したかチェックする。すなわちBがディスク21の最終LBAを超えたかどうかチェックし、超えていれば処理を完了し、超えていなければステップ2002に戻って、処理を繰り返す。

【0016】次に図4で、データ移行手段が図3のデータ移行処理を行っている間に、上位ホスト、すなわち本実施形態ではバックエンドサーバ3からのライト要求があった場合の処理を説明する。この処理は簡単で、ステップ2101でディスク21にデータを書き込み、ステップ2102でテーブル100の該当するLBAのフラグ102を1に更新する。つまり、データ移行処理中にライト処理が行われたLBAについては、ディスク21からのデータ移行は行われない。

【0017】図5で、データ移行手段が図3のデータ移行処理を行っている間に、上位ホスト、すなわち本実施形態ではバックエンドサーバ3からのリード要求があった場合の処理を説明する。ステップ2201で、テーブル100内のリード要求のあったLBAについてフラグ102を参照する。ステップ2202でフラグ102が1かどうかチェックして処理を分岐する。フラグが1の場合には、そのLBAについてはディスク21からのデータ移行が完了しているので、ディスク21からデータを読み出す(ステップ2203)。フラグが0の場合には、そのLBAについてデータ移行が完了していないので、一旦ディスク21からディスク21にデータをコピーする(ステップ2205)。続いてテーブル100のフラグ102を1に更新して(ステップ2206)、ステップ2203以降へ進む。ステップ2204で読み出したデータをバックエンドサーバ3に渡して処理は完了する。

【0018】次に、本実施形態のシステムでの、旧記憶装置サブシステム2から新記憶装置サブシステム2へのデータ移行処理について、システム全体の流れを説明していく。データ移行を行う際、ユーザはバックエンドサーバ3に移行を指示する。バックエンドサーバ3から新記憶装置サブシステム2へのデータ移行処理開始の指示は、インタフェース32を介して新記憶装置サブシステ

ム2に伝えられる。図6はバックエンドサーバ3の処理の流れを説明している。バックエンドサーバ3は移行の指示を受けると、まず、仮想デバイスドライバ31による仮想ディスクの動作を停止する(ステップ1001)。これにより、仮想デバイスドライバ31から旧記憶装置サブシステム2へのアクセスは中断され、仮想デバイスドライバ31はホスト1から仮想ディスクに対するアクセスコマンドを受け付けても、アクセス中止が解除されるまで応答を返さない。次に記憶装置管理プログラム34は新記憶装置サブシステム2に対してデータ移行処理の開始を指示する(ステップ1002)。新記憶装置サブシステム2の行うデータ移行処理については後述する。ステップ1003では、仮想デバイスドライバ31がこれまでホスト1に見せていた仮想デバイスの設定を、ディスク21へのアクセスを行うように変更し、ステップ1004ではステップ1001で中止していたアクセスを再開させる。仮想ディスクのアクセスが再開されると、ステップ1001、ステップ1002の間にホスト1から仮想ディスクに対してアクセスがきていた場合、そのアクセスは全てディスク21に対して実施される。

【0019】また、本実施形態においては、記憶装置サブシステム2とバックエンドサーバ3が直接スイッチにつながった接続形態であったが、図7のように記憶装置サブシステム2がバックエンドサーバ3を介してつながる構成であっても実現は可能である。さらに、新規に増設する記憶装置サブシステム2が図1の例のようにデータ移行手段24をもたないような場合には、図8のようにバックエンドサーバ3側にデータ移行手段24を持たせ、バックエンドサーバでデータ移行処理を行わせることで同様のことが実現できる。また、本実施形態ではバックエンドサーバ3を設けてホストに仮想的なディスクを見せる処理を施しているが、図9のように仮想デバイスドライバ31、記憶装置管理プログラム34、そしてデータ移行手段24をファイバチャネルスイッチ4に持たせるという構成も可能である。本実施形態では、ホストコンピュータに透過的にディスク間のデータ移行ができる例を示したが、さまざまな適用先がある。仮想デバイスドライバが仮想ディスクと実際の記憶装置との対応付けを行えば、ホストコンピュータにとっては実際にデータがどの記憶装置にあってもかまわない。そのため、例えば普段は必要最低限の仮想ディスクを定義しておき、必要になった時点で動的に必要な容量の仮想ディスクを用意できるような記憶装置管理システムや、データのアクセス頻度により、ホストに透過的にデータを低速ディスクから動的に高速ディスクに移動するシステムなどに応用できる。

【0020】

【発明の効果】本発明によれば、ホストコンピュータに対して一切透過的にデータ装置間のデータ移動や、デ

ディスク容量のオンライン拡張など、あらゆるデータ操作が可能となる。

【図面の簡単な説明】

【図1】本発明の実施形態における計算機システムの構成例を示すブロック図である。

【図2】本発明の新記憶装置サブシステムのデータ移行手段の使用するテーブルを示すテーブル構成図である。

【図3】本発明のデータ移行手段が行うデータ移行処理の流れを示すフローチャートである。

【図4】データ移行処理中にライト要求が来たときの、データ移行手段の処理の流れを示すフローチャートである。

【図5】データ移行処理中にリード要求が来たときの、データ移行手段の処理の流れを示すフローチャートである。

【図6】本発明の実施形態における計算機システムにおいて、旧記憶装置サブシステムから新記憶装置サブシステムへのデータ移行処理を行うときの、バックエンドサ

ーバの処理の流れを示すフローチャートである。

【図7】本発明の実施形態を実現する、別の計算機システムの構成例を示すブロック図である。

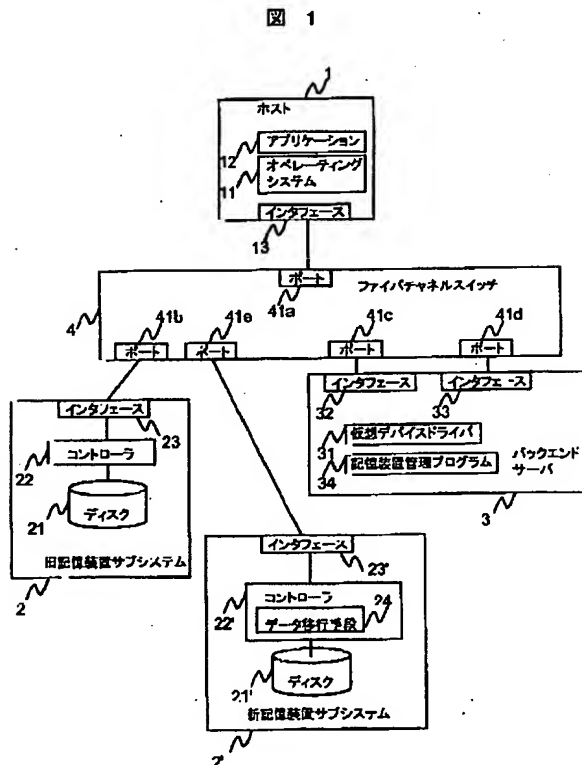
【図8】本発明の実施形態を実現する、別の計算機システムの構成例を示すブロック図である。

【図9】本発明の実施形態を実現する、別の計算機システムの構成例を示すブロック図である。

【符号の説明】

1…ホスト、2…旧記憶装置サブシステム、2'…新記憶装置サブシステム、3…バックエンドサーバ、4…ファイバチャネルスイッチ、11…オペレーティングシステム、12…アプリケーション、13…インタフェース、21…ディスク、22…コントローラ、23…インタフェース、24…データ移行手段、31…仮想デバイスドライバ、32…インタフェース、33…インタフェース、34…記憶装置管理プログラム、41a…ポート、41b…ポート、41c…ポート、41d…ポート、41e…ポート。

【図1】

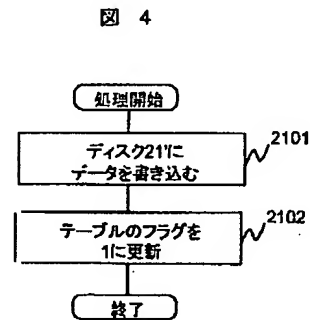


【図2】

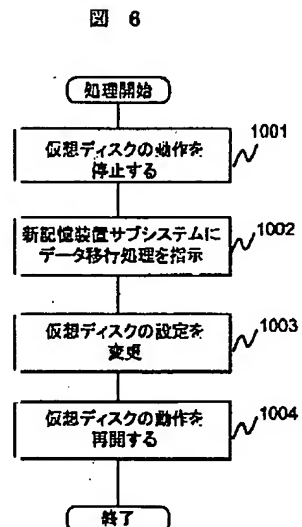
図 2

アドレス	フラグ
LBA0	1
LBA1	1
LBA2	0
LBA3	0
LBA4	1
...	...
LBA _n	0

【図4】

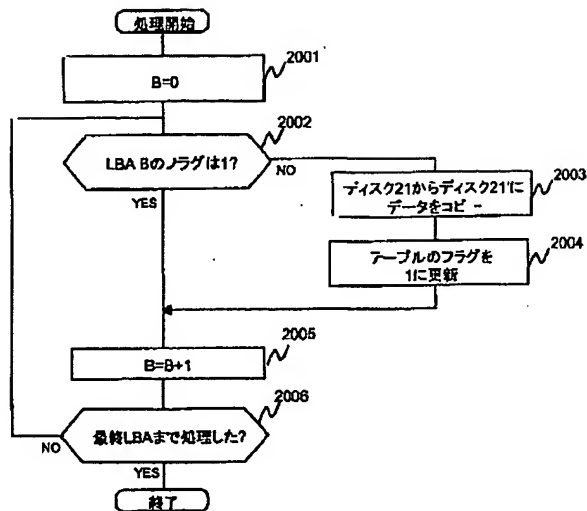


【図6】



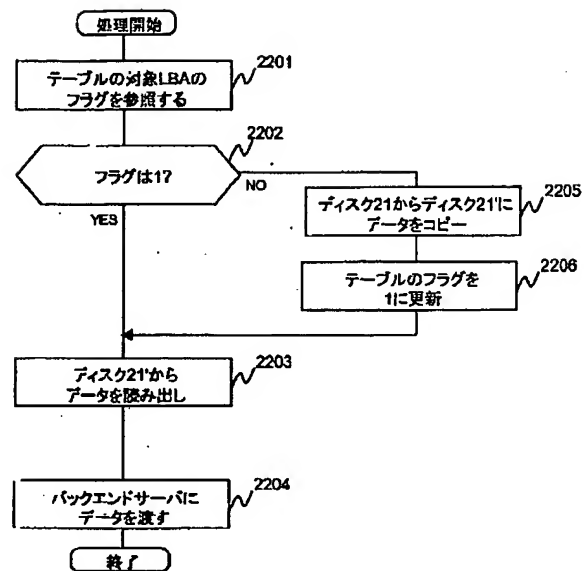
【図3】

図 3



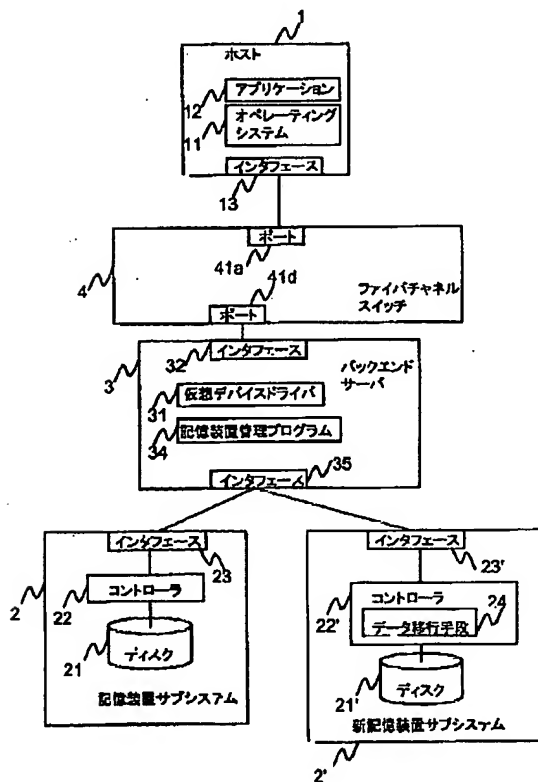
【図5】

図 5



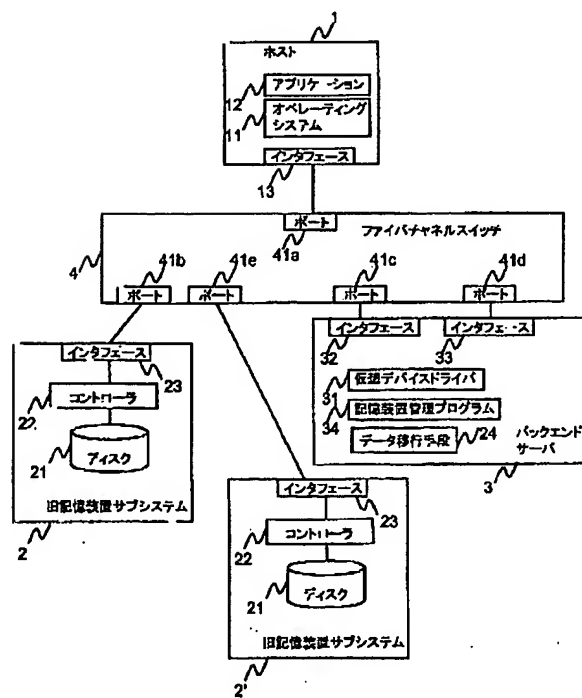
【図7】

図 7



【図8】

図 8



【図9】

図 9

